ABSTRACT

DISSERTATION: Implications of missing data designs with the use of a longitudinal dataset

STUDENT: Melissa Singh

DEGREE: Doctor of Philosophy

COLLEGE: Teachers College

DATE: May 2017

PAGES: 83

This study assessed various methods that account for managing missing data when analyzed with a longitudinal dataset. Moreover, the paper investigated the performance of traditional and modern missing data methods, as a means to providing researchers with a greater understanding of the statistical tools available to them when faced with missing data. This study utilized an archival, publicly available dataset, that consisted of 16,197 participants. To explicate the methods and magnify the generalizability of the results, the overall sample was randomly split into four equally sized subsamples, prior to running any analyses. The results overall, indicated listwise data deletion yielded relatively higher standard errors for the regression coefficients, when compared to the other models, and the results were consistent with past research (e.g., Allison, 2002; Witta, 2000; Baraldi & Enders, 2010). Results from the arithmetic mean imputation and maximum likelihood presented with the most statistically significant predictor variables, but in the case of mean imputation, this may have been due to an underestimation of the standard errors of the coefficients. Multiple imputation with MICE displayed few statistically significant variables across all four subsamples. In regards to all of the missing data approaches, multiple imputation using MICE, displayed the logistic regression parameter estimates with the least extreme standard errors. This does not mean it is uniformly the

best method to use, as it has its weaknesses (e.g., computation limitations with a large number of interactions), but it may be a preferred method for less complex methods.